



Amirkabir University of Technology (Tehran Polytechnic)  
Dept. of Industrial Engineering & Management System

## Data & Information Analysis Course Fall 1400 Final Project

مجموعه داده مربوط به تعداد درخواست جهت اجاره دوچرخه به پیوست آورده شده است. هدف این پروژه، تحلیل داده توصیفی و تجویزی (پیش‌بینی تعداد درخواست‌ها) به کمک مطالب تدریس شده است.



توضیحات ویژگی‌ها به شرح زیر است:

**datetime** - hourly date + timestamp  
**season** - 1 = spring, 2 = summer, 3 = fall, 4 = winter  
**holiday** - whether the day is considered a holiday  
**workingday** - whether the day is neither a weekend nor holiday  
**weather** - 1: Clear, Few clouds, Partly cloudy, Partly cloudy  
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist  
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds  
4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog  
**temp** - temperature in Celsius  
**atemp** - "feels like" temperature in Celsius  
**humidity** - relative humidity  
**windspeed** - wind speed  
**casual** - number of non-registered user rentals initiated  
**registered** - number of registered user rentals initiated  
**count** - number of total rental

مراحل انجام پروژه:

بخش اول: (۰.۲۵ نمره)

• Import کردن کتابخانه‌های مورد نیاز (کتابخانه‌های نظیر Seaborn, Pandas, Numpy و ...)

بخش دوم: (۰.۲۵ نمره)

• بارگذاری و خواندن دیتاست (با استفاده از کتابخانه Pandas)

- اکتشاف دیتاست و بررسی داده‌ها (با استفاده از ابزارهای کتابخانه Pandas مانند describe و ... )
- بخش سوم: (۰.۲۵ نمره)
- شناسایی داده‌های گم‌شده (Missing Value)
- بخش چهارم: (۰.۵ نمره)
- بررسی ارتباط میان ویژگی‌ها و بدست آوردن همبستگی (Covariance) میان ویژگی‌ها و رسم نمودار همبستگی میان ویژگی‌ها و انتخاب ویژگی‌هایی با بیشترین همبستگی با تعداد درخواست دوچرخه (Feature Selection)
- بخش پنجم: (۰.۷۵ نمره)
- انتخاب مدل مناسب (از میان مدل‌های تدریس شده) و آموزش مدل (با استفاده از فایل train.csv) و مقایسه و ارزیابی قدرت پیش‌بینی مدل‌های بدست آمده با داده‌های test (با استفاده از فایل test.csv)

- ✓ پاسخ‌ها را حتما در قالب یک فایل **Jupyter Notebook** یا در محیط **google colab** به همراه خروجی کدها ذخیره و ارسال کنید. در غیر این صورت تصحیح نخواهد شد و بدیهی است که نمره ای نیز تعلق نخواهد گرفت.
- ✓ ارسال فایل گزارش الزامی نبوده ولی باعث فهم بهتر کد شما شده و بسیار توصیه می‌شود.
- ✓ مجموع نمرات امتیازی مرتبط با نرم افزار ۲+ نمره می‌باشد و در نمره نهایی لحاظ خواهد شد.